Assay for detecting and identifying micro-organisms

Field of the invention

The present invention relates to an assay and a method for diagnosing and identifying micro-organisms, and in particular bacteria. The present invention also relates to an assay and a method for detecting micro-organisms, and in particular bacteria, in a sample, and for the discrimination thereof.

More in particular the present invention relates to an assay and a method for the molecular identification of bacteria according to Gram-, genus- species- and strain-specificity based on multigenotypic testing of bacterial DNA from human, animal or environmental samples.

Background

In the medical and veterinary clinical setting, detection and species identification of harmful bacteria infecting biological fluids or tissues is a pre-requisite for appropriate and timely relevant antibiotherapy. Such identification is classically performed by conventional microbiological methods (culture on solid medium or in liquid phase). These conventional methods have however their own limitations.

Culture is always followed by phenotypic identification, which is based on the biochemical features of the bacteria. Usually, the whole process requires 48 to 72 hours to be completed. This period is unfortunately too long, considering the speed of bacterial growth in infected tissues and, for some bacteria, the pathological effects related the toxins that they produce. This time is also too long when bacteria are spread in the environment (aerosol, food or water contamination), where germs are able to infect humans or animals and spread rapidly on a epidemic way from an infected to a healthy body on a very short time. There is therefore a need for the rapid detection and identification of pathogenic bacterial agent(s) involved in human or animal infections or present in the environment.

A stream of studies carried out recently has confirmed that molecular identification is more efficient than phenotypic identification (Bosshard *et al*, 2003; Bosshard *et al*, 2004; Lecouvet *et al*, 2004) and genotypic definition of bacteria species has now become the gold standard (Clarridge, 2004). There is therefore an increasing need for identifying bacterial species with more reliable methods. While obvious in the hospital setting, it is also of interest of the post September 2001 era, where accuracy and speed in identification of deadly bacteria are priorities.

Aside of the time required for routine microbiologic detection, another limiting factor is sometimes the lack of bacterial growth, generating a false-negative microbiologic result. False-negative bacterial cultures are not unusual in the clinical practice, even when clinical and biological signs clearly suggest a florid and active infection (Lecouvet *et al*, 2004). This false-negativity may be due to a low organism burden, non-culturable or slowly growing micro-organisms or, most often, to prior antibiotic therapy (Trampuz *et al*, 2003; Tzanakaki *et al*, 2003). In this case, a false-negative result hampers correct etiological diagnosis regarding the bacterial origin of the infectious disease, and precludes the use of early targeted antibiotherapy. As delayed antibiotherapy may increase the risk of worse clinical outcome (Gutierrez *et al*, 1998; Yu *et al*, 2003, Lecouvet *et al*, 2004), this situation often prompts the use of empiric, broad spectrum and sometimes long-term therapy, and certainly when there is no microbiologic result.

The higher sensitivity, speed and accuracy of DNA amplification by PCR for identification of bacteria is expected to reduce the time to diagnosis, to improve the diagnostic rate, and to allow an early choice of specific antibiotic treatment. Over the last decade, this expectation has fuelled the development of numerous promising DNA assays for detecting and identifying bacteria at the species- or genera-level in human and environmental samples (Jonas *et al*, 2003 ; Palomares *et al*, 2003 ; Poyart *et al*, 2001 ; Xu *et al*, 2002).

These assays remain however restricted to single species and/or genera (Brakstad *et al*, 1992 ; Poyart *et al*, 2001 ; Vannuffel *et al*, 1998). Such restriction has various disadvantages. For instance, in the absence of any indication on the presence of bacterial agents in an environmental sample or in a biological tissue/fluid sample from human or animal origin suspected to be infected but showing no bacterial background due to the presence of a normal bacterial flora, molecular screening methods have to be applied which target the greatest as possible number of potentially pathogenic bacteria including the most feared bacteria (*Staphylococci, Streptococci, Bacillus anthracis, Enterobacteriacea, Neisseria, etc*...) that could be used by bioterrorists. In this case, the use of specific markers or well-defined genera requires multiple and/or repeated testing to confirm or exclude a bacterial diagnosis. Considering the cost of this strategy as well as the limited amount DNA usually available for one sample, this is practically impossible to be performed.

In another example, in samples from tissues showing a bacterial background due the presence of a normal flora, the identification of a well defined panel of pathogenic bacteria recognized as "prior key targets" in the clinical setting considered (e.g. community-acquired pneumonia) remains very difficult.

In view of the above, there is therefore a need for the rapid detection and identification of pathogenic bacterial agent(s) involved in human or animal infections or present in the environment.

There is also a need for identification and diagnostic tools, which allow screening for the presence of pathogenic bacterial agent(s), and to detect and identify these pathogenic bacteria within a bacterial background.

In particular, it is clear that there is a great need in the art for molecular screening/detection and identification assays and methods having a range of specificity that is as wide as possible in order to quickly detect the presence of bacteria (bacterial detection step), while allowing in parallel or subsequently, to identify the present bacterial species, genera and, optionally the strain (bacterial identification step).

In a first aspect, the present invention therefore aims to provide an improved assay for detecting micro-organisms, and in particular bacteria. It is further an aim of the invention to provide an improved assay for diagnosing bacterial infection of a sample and/or tissue.

In another aspect, the present invention also aims to provide an improved assay and method for the identification of micro-organisms. More in particular, the invention aims to identify and provide a series of specific, molecular markers for the detection and/or identification of micro-organisms, and preferably bacteria, in a Gram-, genus- species- and/or strain-specific way.

Summary

The present invention relates to an assay for detecting and identifying one or more micro-organisms in a sample, characterized in that said assay comprises the use of at least two conserved molecular markers. Preferably said micro-organisms are bacteria. In a preferred embodiment, the assay of the present invention is characterized in that it comprises the use of at least one molecular marker that is conserved in Gram-positive bacteria and at least one molecular marker that is conserved in Gram-negative bacteria.

In the prior art, in order to detect the presence of bacteria in samples or tissues, extremely conserved molecular markers are generally used. The most commonly used sequences for detecting bacteria are the sequence of the gene coding for ribosomal DNA (16s rDNA gene) (Klaschik *et al*, 2002) and the 16S-23S intergenic region (Gurtler & Stanisich, 1996). However, ribosomal gene 16S rDNA does not always allow the distinction between species, as illustrated for the *Bacillus* species (La Scola *et al*, 2003). This is a major drawback in the 16S rDNA gene sequence identification method, because in some species, a sequence can be ambiguous since it does not distinguish between two closely related

clinical species but disclosing however a distinct virulence phenotype (for instances, *Escherichia coli* K12 versus *Escherichia coli* O157:H7). This remark applies to the intergenic spacer 16S-23S rDNA as well (Gianinno *et al*, 2003). There is therefore a need to develop a molecular identification system which better discriminate bacteria than the 16S rDNA and the intergenic spacer 16S-23S rDNA.

In accordance with the present invention two series of conserved molecular markers were identified and characterized which are extremely suitable for permitting the detection and genotyping of micro-organisms, and in particular of bacteria, in a Gram-specific way. More in particular these molecular markers comprise on one hand markers preferentially conserved in Gram-positive bacteria and the other hand markers that are preferentially conserved in Gram-negative bacteria. The present invention now allows , by a combined use of these two types of conserved molecular marker sequences, to detect bacteria in a sample and to genotype these bacteria in a gram-specific way as well as in a genera-, species-, and even sometimes, strain-specific manner.

So far, in conventional microbiology, one distinguishes the bacteria according to the structure of their wall (the wall is present in all bacteria except mycoplasms). This structure conditions the color of bacteria after Gram staining (Gram is made of several successive steps including treatment with purple gentian, Lugol's solution, alcohol and fuchsine). The bacteria whose wall is permeable to alcohol lose their purple staining (violet gentian) and coloured in red (fuchsine), defining so what is considered as a Gram-negative bacteria. In Gram-positive bacteria, the wall is primarily made by peptidoglycane. In Gram-negative bacteria, the peptidoglycane layer is thin and the wall has a more complex structure. In the clinical practice, the choice of antibiotherapy relies primarily on Gram stain. Indeed, antibiotics targeting the bacterial wall are much more on Gram-positive bacteria. As already stated above, several clinical studies show that any delay with the initiation of antibiotherapy results in increased mortality and hospital morbidity. Practically, the microbiologic identification (culture) comes too late.

Such approach provides many advantageous compared to conventionally applied detection strategies, wherein no such gram-specificity is involved. The present invention now permits by the use two series of conserved molecular markers to rapidly determine the gram-phenotype of bacteria in a sample and as a consequence to rapidly determine the most suitable antibiotherapy to be applied. This can be substituted to the conventional Gram staining procedure which is far less sensitive.

In a preferred embodiment, the assay of the present invention is further characterized in that the molecular maker that is conserved in Gram-positive bacteria comprises PurA or

PstI. In a more preferred embodiment the molecular maker that is conserved in Gram-positive bacteria is selected from the group comprising the Spy0160 (marker I), Spy1372 (marker II), SpyM3_0902 & SpyM3_0903 (marker III) and Spy1527 (marker IV) marker sequences. In yet another more preferred embodiment, the molecular maker that is conserved in Gram-positive bacteria is selected from the group comprising the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425.

Another preferred embodiment of the invention relates to an assay that is characterized in that the molecular maker that is conserved in Gram-negative bacteria is selected from the group comprising the Ecs0036 (marker V), HI1576 (marker VI), EG10839 and EG11396 (marker VII), and HI0019 (marker VIII) sequences.

In yet another preferred embodiment the molecular maker that is conserved in Gram-negative bacteria is selected from the group comprising the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

Table 1 summarizes sequences used in accordance with the present invention for the detection and identification of Gram-positive and Gram-negative bacteria.

Table 1

Gram-positive bacteria					
Sequence	**Gene**	**Marker**	**Sequences found in Gram-positive bacteria (SEQ ID NOs:)**	**Overlapping sequences found in Gram-negative bacteria (SEQ ID NOs:)**	**Sequences found in other organisms (SEQ ID NO:)**
Spy0160	PurA	I	1-62 ; 326-359	63	
Spy1372	PstI	II	64-107 ; 109-111 ;117-129 ; 137 ; 145-148 ; 360-395 ; 397-399	108 ; 112-116 ; 130-136 ; 138-144 ; 396 ; 400-403	149 (*Cryptococcus neoformans*)
SpyM3_0902 & SpyM3_0903	Hypothetical protein	III	150-180; 404-412		
Spy1527	Hypothetical protein	IV	181-193; 413-425		

Gram-negative bacteria					
Sequence	**Gene**	**Marker**	**Sequences found in Gram-negative bacteria**	**Overlapping sequences found in Gram-positive bacteria**	**Sequences found in other organisms**

Ecs0036	carB	V	194-232 ; 238-239 ; 242-254 ; 431-442	233-237 ; 240-241 ; 255	
HI1576	pgi	VI	256-277; 426-430		
EG10839 & EG11396	sfrB & yigC	VII	278-303; 443-451		
HI0019	yleA	VIII	304-325; 452-461		

The present invention also relates to the use of an assay as defined herein for diagnosing bacterial infection of a sample.

The foregoing and other objects, features and advantages of the invention will become more readily apparent from the following detailed description of preferred embodiments.

Description of the figures

Figure 1 represents the amplification of a molecular marker I (Spy0160 or *pur A)* in Gram-positive bacteria.

Figure 2 represents the amplification of a molecular marker II (Spy1372 or *ptsI*) in Gram-positive bacteria.

Figure 3 represents the amplification of a molecular marker III (SpyM3_0902 & SpyM3_0903) in Gram-positive bacteria.

Figure 4 represents marker I (purA) sequences amplified from different Gram-positive bacteria (SEQ ID NOs 1-62), and from a Gram-negative bacterium (SEQ ID NO: 63)

Figure 5 represents marker II (pstI) sequences amplified from Gram-positive bacteria (SEQ ID NOs: 64-107; SEQ ID NOs: 109-111, SEQ ID NOs: 117-129, SEQ ID NO: 137, SEQ ID NOs 145-148), from some Gram-negative bacteria (SEQ ID NOs 108, 112-116, 130-136, 138-144) and from the fungi *Cryptococcus neoformans* (SEQ ID NO: 149).

Figure 6 represents marker III (SpyM_0902 &SpyM_0903) sequences amplified from Gram-positive bacteria (SEQ ID NOs 150-180).

Figure 7 represents marker IV (putative GTP-binding factor plus 160 nt downstream this ORF) sequences amplified from Gram-positive bacteria (SEQ ID NOs 181-193)

Figure 8 represents the amplification of a molecular marker V (Ecs0036 or carB) in Gram-negative bacteria.

Figure 9 represents sequences amplified with molecular marker V (carB) from various Gram-negative bacteria (SEQ ID NOs 194-232, 238-239, 242-254) and from various Gram-positive bacteria (SEQ ID NOs 233-237, 240-241, 255)

Figure 10 represents the amplification of a molecular marker VI (HI1576 or pgi) in Gram-negative bacteria.

Figure 11 represents sequences amplified with molecular marker VI (HI1576 or pgi) from various Gram-negative bacteria (SEQ ID NOs 256-277).

Figure 12 represents sequences amplified with molecular marker VII (EG10839 & EG11396 or sfrB & yigC) in Gram-negative (SEQ ID NOs 278-303).

Figure 13 represents sequences amplified with molecular marker VIII (HI0019 or hypothetic yleA protein) in Gram-negative bacteria (SEQ ID NOs 304-325).

Figure 14 represents marker I (Spy0160 or purA) sequences amplified from different Gram-positive bacteria (SEQ ID NOs 326-359).

Figure 15 represents marker II (Spy1372 or pstI) sequences amplified from Gram-positive bacteria (SEQ ID NOs: 360-395; SEQ ID NOs: 397-399), and some Gram-negative bacteria (SEQ ID NOs 396, 400-403).

Figure 16 represents marker III (SpyM_0902 &SpyM_0903) sequences amplified from Gram-positive bacteria (SEQ ID NOs 404-412).

Figure 17 represents marker IV (Spy1527, a putative GTP-binding factor plus 160 nt downstream) sequences amplified from Gram-positive bacteria (SEQ ID NOs 413-425).

Figure 18 represents sequences amplified with molecular marker VI (HI1576 or pgi) from various Gram-negative bacteria (SEQ ID NOs 426-430).

Figure 19 represents sequences amplified with molecular marker V (Ecs0036 or carB) from various Gram-negative bacteria (SEQ ID NOs 431-442).

Figure 20 represents sequences amplified with molecular marker VII (EG10839 & EG11396 or sfrB & yigC) in Gram-negative (SEQ ID NOs 443-451).

Figure 21 represents sequences amplified with molecular marker VIII (HI0019, hypothetic yleA protein) in Gram-negative bacteria (SEQ ID NOs 452-461).

Detailed description of the invention

The following definitions serve to illustrate the terms and expressions used in the different embodiments of the present invention as set out below.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated.

The term "probe" or "nucleic acid probe" refers to single stranded sequence-specific oligonucleotides which have a base sequence which is sufficiently complementary to hybridize to the target base sequence to be detected.

The term "primer" refers to a single stranded DNA oligonucleotide sequence capable of acting as a point of initiation for synthesis of a primer, extension product which is complementary to the nucleic acid strand to be copied. The length and the sequence of the primer must be such that they allow to prime the synthesis of the extension products. Preferably the primer is about 5-50 nucleotides long. Specific length and sequence will depend on the complexity of the required DNA or RNA targets, as well as on the conditions of primer use such as temperature and ionic strength.

The term "target" refers to nucleic acid molecules originating from a biological sample which have a base sequence complementary to the nucleic acid probe of the invention. The target nucleic acid can be single-or double-stranded DNA (if appropriate, obtained following amplification) and contains a sequence which has at least partial complementarity with at least one probe oligonucleotide.

The phrase "a (biological) sample" refers to a specimen such as a clinical sample (pus, sputum, blood, urine, etc.) of human or animal, an environmental sample, bacterial colonies, contaminated or pure cultures, purified nucleic acid, etc. in which the target sequence of interest is sought.

The term "polynucleic acid" corresponds to either double- stranded or single-stranded cDNA or genomic DNA, containing at least 10, 20, 30, 40 or 50 contiguous nucleotides.

A polynucleic acid which is smaller than 100 nucleotides in length is often also referred to as an oligonucleotide. Single stranded polynucleic acid sequences are always represented in the present invention from the 5' end to the 3' end. By "oligonucleotide" is meant a nucleotide polymer generally about 10 to about 100 nucleotides in length, but which may be greater than 100 or shorter than 10 nucleotides in length.

The term "homologous" is synonymous for identical and means that polynucleic acids which are said to be e. g. 90% homologous show 90% identical base pairs in the same position upon alignment of the sequences.

"Hybridization" involves the annealing of a complementary sequence to the target nucleic acid (the sequence to be detected). The ability of two polymers of nucleic acid containing complementary sequences to find each other and anneal through base pairing interaction is a well-recognized phenomenon.

The term "stringency" indicates one used to describe the temperature and solvent composition existing during hybridization and the subsequent processing steps. Under high

stringency conditions only highly complementary nucleic acid hybrids will form; hybrids without a sufficient degree of complementarity will not form. Accordingly, the stringency of the assay conditions determines the amount of complementarity needed between two nucleic acid strands forming a hybrid. Stringency is chosen to maximize the difference in stability between the hybrid formed with the target and the non-target nucleic acid.

By "complementary" is meant a property conferred by the base sequence of a single strand of DNA which may form a hybrid or double stranded DNA: DNA, through hydrogen bonding between Watson-Crick base pairs on the respective strands. Adenine (A) usually complements thymine (T), while guanine (G) usually complements cytosine (C).

By "hybrid" is meant the complex formed between two single stranded nucleic acid sequences by Watson-Crick base pairings or non-canonical base pairings between the complementary bases.

Molecular Marker sequences

In a first aspect, the present invention provides conserved molecular markers for the detection and/or identification of one or more micro-organisms, and preferably bacteria. More in particular, the present invention provides two series of conserved molecular markers which are extremely suitable for permitting the detection and genotyping of micro-organisms, and in particular of bacteria, in a Gram-specific way.

The term "*molecular marker*" and "*molecular marker sequence*" are used herein as synonyms. These terms refer to isolated and purified nucleic acid (DNA) molecules. The term "*conserved molecular marker*" as used herein refers to a coding or non coding DNA sequence, which can be found in the genome of various bacterial species, showing a sequence identity with an original sequence which is superior to or equal to 50%, and preferably superior to or equal to 65%, and more preferably superior to or equal to 80%.

According to the present invention, two series of conserved genetic markers were characterized: one preferentially conserved in Gram-positive bacteria and the other preferentially conserved in Gram-negative bacteria.

In a preferred embodiment, the molecular markers that are conserved in Gram-positive bacteria comprise Spy0160 (PurA) or Spy1372 (PstI). More preferably the markers that are conserved in Gram-positive bacteria are selected from the group comprising Spy0160 Spy1372, SpyM3_0902 & SpyM3_0903, and Spy1527 marker sequences

In another preferred embodiment, the molecular markers that are conserved in Gram-positive bacteria are any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425.

In yet another preferred embodiment, the molecular markers that are conserved in Gram-negative bacteria are selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396, and HI0019.

In another preferred embodiment, the molecular markers that are conserved in Gram-negative bacteria are any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.However, it should be clear from the present invention that the present invention is not limited to the molecular marker sequences conserved in Gram-positive and in Gram-positive bacteria as described herein. Other conserved molecular marker sequences that can be characterized and identified for various other Gram- positive bacteria and other Gram-negative bacteria including according to the invention are considered to be included in the present application as well.

In another embodiment, the invention relates to the use of at least two conserved molecular markers for detecting bacteria in a sample.

In a preferred embodiment, the invention relates to the use of at least two conserved molecular markers for detecting and genotyping a bacterium on the basis of the Gram phenotype in a sample. Preferably, the invention relates to the use of at least one molecular marker that is conserved in Gram-positive bacteria and at least one molecular marker that is conserved in Gram-negative bacteria for detecting and genotyping a bacterium. In particularly preferred embodiment, the invention relates to the use of at least one molecular marker that is conserved in Gram-positive bacteria selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425, and at least one molecular marker that is conserved in Gram-negative bacteria and that is selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396, HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

The present invention thus provides for highly conserved molecular markers that can be used for detecting the molecular presence of micro-organisms, and in particular of bacteria, in samples and/or tissues, including in cultured samples which give a false-negative result using conventional detection techniques. The present conserved markers can also advantageously be used for detecting the molecular presence of micro-organisms, and in particular of bacteria, in samples from tissues showing bacterial background. In the latter case, the conserved molecular markers are preferably used in combination with specific

primers or probes that directly target a pre-defined panel of bacteria of interest and that exclude the "background flora". A suitable pre-defined panel of bacteria of interest may, for instance, include bacteria involved in community-acquired pneumonia, such as but is not limited to *Haemophilus influenzae, Legionella species, Staphylococcus aureus, Moraxella catarrhalis, Gram-negative enteric bacteria.*

It is further noted that the molecular identification of Gram phenotype is based on partially overlapping Gram-positive and Gram-negative markers. It must born in mind that, unlike in the present invention, another conserved marker (16S) shows an extensive overlap between Gram-positive and Gram-negative bacteria. In the present case, using concomitantly both series of partially overlapping markers in a combined way makes it possible to cover a much broader spectrum of bacterial pathogens while defining also precisely the Gram phenotype of those pathogens. The strategy relies upon the molecular detection of gene preferentially present in Gram-positive or Gram-negative bacterial. Each series of markers allows therefore improving overall detection in their respective group (either Gram-positive bacteria for preferentially Gram-positive markers, or Gram-negative bacteria for preferentially Gram-negative markers). Considering the somehow overlapping specificity for both groups, (overlap within the Gram-positive specificity for Gram-negative markers and overlap within the Gram-positive specificity for Gram-negative markers), the power of the molecular discrimination is even increased for some bacteria targeted by both groups of markers. This combined strategy overcomes the potential lack of specificity obtained when using one single marker towards some species, as is for instance the case when using a 16S marker.

In addition, the use of different markers which are mapped on different loci in the bacteria also improves the quality of the diagnosis in that it can more easily circumvent false positive reactions due to accidental PCR contamination hampering the use of one particular marker.

Primers and probes derived from conserved molecular markers

In another embodiment, the invention relates to a primer pair (forward and reverse primers) suitable for amplifying a molecular marker that is conserved in Gram-positive bacteria. More preferably, the invention relates to a primer pair suitable for amplifying any of the conserved molecular marker sequences that are conserved in Gram-positive bacteria as defined herein, and that are preferably selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID

NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425.

In another embodiment, the invention relates to a primer pair (forward and reverse primers) suitable for amplifying a molecular marker that is conserved in Gram-negative bacteria. More preferably, the invention relates to a primer pair suitable for amplifying any of the conserved molecular marker sequences that are conserved in Gram-negative bacteria as defined herein, and that are preferably selected from the group comprising Ecs0036, HI1576,, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

The primers of the present invention include at least 15-mer oligonucleotide and are preferably 70%, 80%, 90% or more than 95% homologous to the exact complement of the target sequence to be amplified. Those primers are about 15 to 50 nucleotides long, and preferably about 15 to 35 nucleotides long. Of course, primers consisting of more than 50 nucleotides can be used.

The present invention also relates to a nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria. More preferably, the invention relates to a nucleic acid probe capable of hybridizing any of the molecular marker sequences that are conserved in Gram-positive bacteria as defined herein, and that are preferably selected from the group comprising Spy0160 (PurA), Spy1372 (PstI), SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425.

In another embodiment, the present invention also relates to a nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria. More preferably, the invention relates to a nucleic acid probe capable of hybridizing any of the molecular marker sequences that are conserved in Gram-negative bacteria as defined herein, and that are preferably selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

The probe of the present invention preferably includes at least 15-mer oligonucleotide and are preferably 70%, 80%, 90% or more than 95% homologous to the exact complement of the target sequence to be detected. Those probes are preferably about 15 to 50 nucleotides long.The primers and probes of the invention can be used, for diagnostic purposes, in investigating the presence or the absence of a target nucleic acid in a biological

sample, according to all the known hybridization techniques such as for instance dot blot, slot blot, hybridization on arrays including nanotools, real-time PCR, etc...

The probes of the invention will preferably hybridize specifically to one or more of the above-mentioned molecular marker sequences.

The primers of the invention may amplify specifically one or more of the above-mentioned marker sequences. The design of specifically hybridising probes is within the skilled person's knowledge. Also the design of primers which can specifically amplify certain sequences or molecular markers is within the skilled person's knowledge.

The nucleic acid probes of this invention can be included in a composition or kit which can be used to rapidly determine the presence or absence of pathogenic species of interest (see below).

Compositions

In another embodiment, the invention relates to a composition. In a preferred embodiment, the invention relates to a composition comprising at least one primer pair (forward and reverse primers) suitable for amplifying a conserved molecular marker that is conserved in Gram-positive bacteria and at least one primer pair (forward and reverse primers) suitable for amplifying a conserved molecular marker that is conserved in Gram-negative bacteria.

Preferably, the composition comprises at least one primer pair suitable for amplifying any of the molecular marker sequences that are conserved in Gram-positive bacteria and that are selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425, and at least one primer pair suitable for amplifying any of the molecular marker sequences that are conserved in Gram-negative bacteria and that are selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

In yet another embodiment, the invention relates to a composition comprising at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria, and at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria. Preferably, the composition comprises at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria selected from the group comprising Spy0160, Spy1372,

SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425, and at least one at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

By "composition", it is meant that primers or probes complementary to bacterial DNA may be in a pure state or in combination with other primers or probes. In addition, the primers or probes may be in combination with salts or buffers, and may be in a dried state, in an alcohol solution as a precipitate, or in an aqueous solution.

Kits

In yet another embodiment, the invention relates to a kit for detecting and identifying one or more micro-organisms, preferably bacteria, in a sample, which comprises:

a) a composition comprising at least one primer pair (forward and reverse primers) suitable for amplifying a conserved molecular marker that is conserved in Gram-positive bacteria and at least one primer pair (forward and reverse primers) suitable for amplifying a conserved molecular marker that is conserved in Gram-negative bacteria; for amplifying polynucleic acids in said sample,

b) a composition comprising at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria, and at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria,

c) a buffer enabling hybridization reaction between the probes contained in said composition and the polynucleic acids present in said sample or amplified products therefrom or components necessary for producing the buffer,

d) a solution for washing hybrids formed under the appropriate wash conditions or components necessary for producing the solution, and

e) optionally a means for detection of said hybrids.

A kit according to the invention preferably includes all components necessary to assay for the presence of bacteria. In the universal concept, the kit includes a stable preparation of labeled probes, hybridization solution in either dry or liquid form for the hybridization of target and probe polynucleotides, as well as a solution for washing and

removing undesireable and nonduplexed polynucleotides, a substrate for detecting the labeled duplex, and optionally an instrument for the detection of the label.

In a preferred embodiment, the present kit comprises a composition which comprises at least one primer pair suitable for amplifying any of the molecular marker sequences that are conserved in Gram-positive bacteria selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425, and at least one primer pair suitable for amplifying any of the molecular marker sequences that are conserved in Gram-negative bacteria selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

In yet another preferred embodiment, the present kit comprises a composition which comprises at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425, and at least one at least one nucleic acid probe capable of hybridising to a molecular marker that is conserved in Gram-negative bacteria selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

In yet another preferred embodiment, the kit according to the present invention further comprises one or more genus-, species and/or strain-specific nucleic acid probes capable of hybridizing to a genus-, species and/or strain-specific bacterial polynucleotide sequence.

DNA chip

In another preferred embodiment, the present invention provides a DNA chip in which nucleic acid probes are immobilized on a solid support. The invention relates to the manufacturing of a solid support (array –DNA chip) on which several sets of nucleic acid probes are covalently bound or directly synthesized.

In a preferred embodiment, the invention relates to a DNA chip in which at least one of nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria, and at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria, is immobilized on a solid support.

Preferably, the DNA chip comprises at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, and Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425, and at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461 immobilized on a solid support.

In yet another preferred embodiment, the DNA chip according to the present invention further comprises one or more genus-, species and/or strain-specific nucleic acid probes capable of hybridizing to a genus-, species and/or strain-specific bacterial polynucleotide sequence.

The DNA chip which is formed by arranging DNA fragments of variety of base sequences on the surface of a narrow substrate in high density is used in finding out the information on DNA of an unknown sample by hybridization between an immobilized DNA and unknown DNA sample complementary thereto. Examples of the solid carrier on which the probe oligonucleotides are fixed include inorganic materials such as glass and silicon and polymeric materials such as acryl, polyethylene terephtalate (PET), polystyrene, polycarbonate and polypropylene. The surface of the solid substrate can be flat or have a multiple of hole. The probes are immobilized on the substrate by covalent bond of either 3'end or 5'end. The immobilization can be achieved by conventional techniques, for example, using electrostatic force, binding between aldehyde coated slide and amine group attached on synthetic oligomeric phase or spotting on amine coated slide, L- lysine coated slide or nitrocellulose coated slide. The immobilization and the arrangement of various probes onto the solid substrate are carried out by pin microarray, inkjet, photolithography, electric array, etc.

The term "DNA chip" as used herein, is to be understood in its broadest sense, i.e. including nanochips or nanotools that are designed to recognize a specific pattern of nucleic acids through hybridization.

Assay

In another embodiment, the invention relates to an assay for detecting and identifying one or more micro-organisms, preferably bacteria, in a sample, characterized in that said assay comprises the use of at least two conserved molecular markers, and preferably

comprises the use of at least one molecular marker that is conserved in Gram-positive bacteria and at least one molecular marker that is conserved in Gram-negative bacteria.

In the prior art, ultimate molecular species identification results classically from sequence analysis of an amplification product and the comparison of this sequence with those which are available in public DNA database (for instance, GeneBank...). The sequence analysis requires nearly 24 hours to complete the various analytical steps: amplicon purification, cycle sequencing, reading and interpretation of the results.

The present invention provides a strategy which permits to significantly reduce the time requested for genera, species, and optionally strain, identification of bacteria in a sample compared to classical identification strategies as described above. More in particular, the present strategy preferably consists in amplifying a set of conserved genetic markers and either to hybridize produced amplicons on specific capture probes covalently bound on an array or, alternatively, to hybridize a specific probe during the amplification step (e.g. real-time PCR with Taqman or molecular Beacon probes). The result of the identification will include information regarding the Gram phenotype of bacteria present in the sample of interest (or a combination of both Gram phenotype in case of mixed infections with Gram-positive and Gram-negative), as well as information on genera and species to which they belong. The final results will integrate all the hybridization signals generated by the selected markers.

In a preferred embodiment, the method for detecting and identifying one or more micro-organisms, preferably bacteria, in a sample may comprise the following steps:

a) If appropriate isolating and/or concentrating the DNA present in said sample;
b) amplifying said DNA with
    - at least one pair of (forward and reverse) primers suitable for amplifying a molecular marker that is conserved in Gram-positive bacteria, and
    - at least one pair of (forward and reverse) primers suitable for amplifying a molecular marker that is conserved in Gram-negative bacteria,
c) hybridizing the amplified DNA fragments obtained in step b) with genus-, and/or species-, and/or strain-specific primers or nucleic acid probes
d) detecting the hybrids formed in step c) and
e) identifying micro-organisms in said sample from the differential hybridization signals obtained in step d)

The present invention allows to detect the presence of bacteria in human, animal and/or environmental samples, and, at the same time, to identify those bacteria, including, highly pathogenic ones. Such detection and identification system is based on the pattern of

hybridization of several combined DNA fragments. Identification relies upon concomitant signals generated by a panel of unrelated markers. The system provides discrimination based on the Gram-phenotype as well as genus- and species-specificity.

To provide nucleic acid substrates for use in the detection and identification of micro-organisms in clinical samples using the present assay, nucleic acid, preferably DNA, is extracted from the sample. The nucleic acid may be extracted from a variety of clinical samples, or environmental samples, using a variety of standard techniques or commercially available kits.

A second step is an amplification of the desired DNA region of the target DNA by PCR. Examples of the PCR include most typical PCR using the same amounts of forward and reverse primers, multiplex PCR in which a multiple of target DNAs can be amplified at once by adding various primers simultaneously, ligase chain reaction (LCR) in which target DNA is amplified using specific 4 primers and ligase and the amount of fluorescence is measured by ELISA (Enzyme Linked Immunosorbent Assay), and the other PCR such as Hot Start PCR, Nest-PCR, DOP-PCR (degenerate oligonucleotide primer PCR), RT-PCR (reverse transcription PCR), Semi-quantitative RT-PCR, Real time PCR, RACE (rapid amplification of cDNA ends), Competitive PCR, STR (short tandem repeats), SSCP (single strand conformation polymorphism), DDRT-PCR (differential display reverse transcriptase), etc.

A further step comprises the hybridization of the amplified DNA fragments obtained in step b) with specific primers or nucleic acid probes. The choice of the applied hybridization technique should not be considered as limitative for the present invention. Nanotools can also be designed to recognize a specific nucleic acid pattern with or without PCR amplification.

In one embodiment, step c) may include an identification by contacting the amplified DNA fragments obtained in step b) with a composition comprising at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-positive bacteria, and at least one nucleic acid probe capable of hybridizing to a molecular marker that is conserved in Gram-negative bacteria. In such case, steps b) and c) of the above described method are performed subsequently. For instance, identification may be performed using a reverse hybridization procedure (dot blot, slot blot, hybridization on micro-, macro- or nano-arrays, etc....) In such case, hybridization of marked amplicons is performed on Gram-, genera- and species-specific nucleic acid probes bound covalently on the array (e.g. micro- or macro-arrays). While the result is visual, the reading can also be automated, facilitating therefore the use in the clinical practice.

In another embodiment, step c) may include the use of real-time PCR with specific probes (multiplex approach) which allows to have a specific result in a couple of hours. Simplex or multiplex PCR techniques in real-time include the use of specific acid probes on a DNA target during the PCR step (e.g. Taqman probes, molecular beacons or MGB (Minor Groove Binding) probes, etc...). In such case, steps b) and c) of the above described method are performed simultaneously. A significant advantage of using a real-time PCR technique is the increased speed, e.g. only a couple of hours are required for obtained final results. This is mainly due to the reduced cycle times, removal of separate post-PCR detection procedures, and the use of sensitive fluorescence detection equipment, allowing earlier amplicon detection. Another advantage of a real-time PCR technique is that it is a single tube procedure with the reading during the PCR and no manipulation required at the end of the procedure. This single tube procedure prevents therefore the risk of molecular contamination existing with other PCR procedures, including micro- and macro-array techniques. Preferably a nested real-time PCR strategy is used to increase the sensitivity of the method with a detection limit of one DNA copy in tissue samples.

The formed hybrids can be quantified by labeling the target with a fluorescence or radioactive isotope in accordance to conventional methods. The labeling may be carried out by the use of labeled primers or the use of labeled nucleotides incorporated during the polymerase step of the amplification.

According to a preferred embodiment of the present invention, the primers suitable for amplifying a molecular marker that is conserved in Gram-positive bacteria used in step b) of the above mentioned method preferably comprise primers suitable for amplifying a molecular marker selected from the group comprising Spy0160, Spy1372, SpyM3_0902 & SpyM3_0903, Spy1527, or any of the sequences with SEQ ID NOs 1-62, 64-107, 109-111, 117-129, 137, 145-148, 150-193, 233-237, 240-241, 255, 326-395, 397-399, 404-425.

According to another preferred embodiment of the present invention, the primers suitable for amplifying a molecular marker that is conserved in Gram-negative bacteria used in step b) of the above mentioned method preferably comprise primers suitable for amplifying a molecular marker selected from the group comprising Ecs0036, HI1576, EG10839 & EG11396 and HI0019, or any of the sequences with SEQ ID NOs 63, 108, 112-116, 130-136, 138-144, 194-232, 238-239, 242-254, 256-325, 396, 400-403, 426-461.

The following examples are meant to illustrate the present invention. **Examples 1-3** present and describe molecular marker sequences that have been characterized on a wide panel of clinical and reference bacterial genera species and bacterial strains and that allow the discrimination between Gram-positive and Gram-negative bacteria as well as genera and

species identification. **Example 4** illustrates the use of molecular markers according to the present invention for the detection and identification of Bacillus bacteria and for the discrimination between various Bacillus strains of the same species. **Example 5** illustrates nucleic acid probes according to the present invention.

Examples

Example 1 Characterization of molecular markers of genes enabling the identification of Gram-positive bacteria

The present example aims to test the Gram-positive specificity of several markers. The list of Gram-positive bacteria used in the present example (study of markers specificity) is given in **table 2A.**

Table 2A Gram-positive bacteria

genus	species	strain reference
Bacillus	*anthracis*	1978
Bacillus	*anthracis*	STERNE
Bacillus	*anthracis*	BUTARE
Bacillus	*anthracis*	1655H85
Bacillus	*anthracis*	CODA - CERVA
Bacillus	*anthracis*	2054H82
Bacillus	*cereus*	ATCC10987
Bacillus	*cereus*	ATCC 14579
Bacillus	*pumilus*	Clinical
Bacillus	*species*	Clinical
Bacillus	*thuringiensis kurstaki*	T03A016 HD_1
Bacillus	*thuringiensis israelensis*	4Q2_72
Bacillus	*mycoides*	MYC003
Bacillus	*mycoides*	NRS306
Bacillus	*weihenstephanensis*	WSBC10204
Bacillus	*halodurans*	DSMZ 497
Bacillus	*firmus*	DSMZ 12643
Bacillus	*megatherium*	DSMZ 1324
Bacillus	*pseudomycoïdes*	DSMZ 12442
Clostridium	*difficile*	DSMZ 1296
Clostridium	*perfingens*	DSMZ 756
Enterococcus	*faecium*	DSMZ 6177
Enterococcus	*faecalis*	DSMZ 2570
Enterococcus	*flavescens*	DSMZ 7370
Enterococcus	*durans*	DSMZ 20633
Enterococcus	*casseliflavus*	DSMZ 20680
Enterococcus	*gallinarum*	DSMZ 20628
Enterococcus	*hirae*	DSMZ 20160
Enterococcus	*raffinosus*	DSMZ 75633

Enterococcus	*avium*	DSMZ 20679
Enterococcus	*villorum*	CODA - CERVA
Lactococcus	*lactis*	DSMZ 20481
Listeria	*monocytogenes*	DSMZ 20600
Listeria	*innocua*	DSMZ 20649
Staphylococcus	*aureus*	ATCC 35884
Staphylococcus	*epidermidis*	ATCC 14990
Staphylococcus	*hominis*	ATCC 27844
Staphylococcus	*haemolyticus*	ATCC 29970
Staphylococcus	*saprophyticus*	ATCC 15305
Staphylococcus	*xylosus*	ATCC 35663
Staphylococcus	*simulans*	ATCC 27848
Staphylococcus	*cohnii cohnii*	ATCC 35662
Staphylococcus	*capitis capitis*	ATCC 27840
Staphylococcus	*sciuri*	ATCC 29062
Staphylococcus	*warneri*	ATCC 27836
Staphylococcus	*lugdunensis*	ATCC 43809
Staphylococcus	*gallinarum*	ATCC C3572
Staphylococcus	*schleiferi schleiferi*	ATCC 43808
Staphylococcus	*capitis ureolyticus*	ATCC 49326
Staphylococcus	*cohnii urealyticum*	ATCC 49330
Staphylococcus	*auricularis*	ATCC 33753
Staphylococcus	*caseolyticus*	ATCC 13548
Staphylococcus	*intermedius*	ATCC 29663
Streptococcus	*pyogenes*	DSMZ 20565
Streptococcus	*agalactiae*	DSMZ 2134
Streptococcus	*pneumoniae*	DSMZ 20566
Streptococcus	*oralis*	DSMZ 20627
Streptococcus	*sanguinis*	DSMZ 20567
Streptococcus	*mitis*	DSMZ 12643
Streptococcus	*gordonii*	DSMZ 6777
Streptococcus	*canis*	DSMZ 20386
Streptococcus	*mutans*	DSMZ 20523
Streptococcus	*subspecies*	Clinical
Streptococcus	*bovis*	DSMZ 20480
Streptococcus	*thermophilus*	DSMZ 20617
Streptococcus	*suis*	DSMZ 9682

The following sequences have been characterized and used for multigenotypic identification of Gram-positive bacteria.

The **Spy0160 sequence** (marker I) from *Streptococcus pyogenes* (accession number: AE006485.1; position 3201 to 4030) is part of an open reading frame homologous for the gene purA. The purA protein plays an important role in the *novo* bacterial synthesis of purines. It catalyses the synthesis of adenylosuccinate starting from inosine monophosphate (IMP) and of aspartate, and using energy provided by the GTP. The first nucleotides alignments performed suggested the existence of conserved sequences similar to the

Spy0160 sequence in a few Gram-positive bacteria. These findings have been extended to a wide panel of bacteria that were available in our DNA bank and confirmed that this marker was present in almost all Gram-positive bacteria and was very helpful in discrimination of closely related Gram-positive species and in many cases could allow discrimination between strains of the same species

A second marker is the **Spy1372 sequence** (Marker II) from *Streptococcus pyogenes* (accession n° AE004092, position 1139277 to 1141010). The corresponding gene probably encodes an enzyme involved in the transport of sugar in bacteria. Indeed, this gene is homologous to gene ptsI of *Staphylococcus aureus* coding for a phosphoenol pyruvate phosphatase (accession n° NC_002758, from position 1137273 to 1138991). This gene is part of PTS operon (phosphotransferase system) including several genes coding for proteins involved in importation of sugar by bacteria (Plumbridge *et al*, 2002). The product of ptsI gene is a protein called Enzyme I, which may be phosphorylated by phosphoenol pyruvate. Phosphorylated Enzyme I can give its phosphate group to another protein of the PTS group through a cascade which leads to the entry of glucose in the bacterial periplasm (Stentz *et al*, 1997).

A third marker is the **SpyM3_0902 - SpyM3_0903 sequence** (Marker III) from *Streptococcus pyogenes* MGAS315 (accession n° AE014154, from position 40670 to 41160). This sequence is located downstream the gene encoding alpha-helicase and corresponds to the open reading frame of a hypothetical protein.

A fourth marker is the **Spy1527 sequence** (Marker IV) from *Streptococcus pyogenes* from position 1201 to 2464, including nucleotides 2465 to 2625 (accession n° AE006586). The Spy1527 sequence corresponds to the gene *typA*, coding for a putative GTP-binding protein (GTP-BP). The fragment from position 2465 to 2625 does not correspond to an open reading frame, but is a non-coding sequence.

A first analysis of some available complete bacterial genomes suggests that homologous sequences were present in most of those bacteria.

A further analysis has been performed on reference strains and on several hundreds clinical strains provided by Belgian hospitals. The conservation of targets of interest (purA and ptsI (i.e. Marker I and II) has been confirmed in the genome of all these reference and clinical strains. This analysis confirmed the very little genomic variability of these sequences within a species of interest. This feature is crucial to allow the use of these marker sequences in a strategy of multigenotypic identification of Gram-positive bacteria.

Example 2 Characterization of molecular markers of genes enabling the identification of Gram-negative bacteria

The present example aims to test the Gram-negative specificity of several markers. The list of Gram-positive bacteria used in the present example (study of markers specificity) is given in **table 2B.**

Table 2B Gram-negative bacteria

genus	species	strain reference
Acinetobacter	*baumanii*	ATCC 19606
Acinetobacter	*calcoaceticus*	DSMZ 1139D
Bordetella	*parapertussis*	Clinical
Bordetella	*bronchiseptica*	Clinical
Bordetella	*pertusis*	Clinical
Brucella	*melitensis biovar 1*	16M
Brucella	*melitensis biovar 2*	63/9
Brucella	*abortus biovar 1*	544
Brucella	*abortus biovar 2*	86/8/59
Brucella	*canis*	RM6/66
Brucella	*ovis*	63/290
Brucella	*suis biovar 1*	1330
Brucella	*suis biovar 2*	686
Burkholderia	*cepacia*	ATCC 17770
Citrobacter	*freundii*	DSMZ 30039
Cryptococcus	*neoformans*	DSMZ 70219
Enterobacter	*cloaceae*	ATCC 13047
Enterobacter	*aerogenes*	DSMZ 13048
Escherishia	*coli O157:H7*	DSMZ 8579
Escherishia	*coli K12*	DSMZ 6367
Francisella	*tularensis*	SVA / T7
Haemophilus	*influenzae*	DSMZ 9999
Haemophilus	*ducrei*	-
Klebsiella	*pneumoniae*	ATCC 13883
Klebsiella	*oxytica*	ATCC 43863
Legionella	*pneumophilia*	DSMZ 7513
Moraxella	*catarrhalis*	DSMZ 11994
Morganella	*morganii*	ATCC 25830
Neisseria	*meningitidis groupe C*	ISP ???
Neisseria	*meningitidis groupe B*	clinical
Pasteurella	*multocida*	-
Proteus	*mirabilis*	ATCC 29906
Proteus	*vulgaris*	ATCC 13315
Providencia	*stuartii*	ATCC 29914
Pseudomonas	*aeruginosa*	DSMZ 50071

Pseudomonas	*putida*	ATCC 12633
Pseudomonas	*syringae*	ATCC 39254
Salmonella	*enteritidis*	Clinical
Salmonella	*enterica hadar*	Clinical
Salmonella	*enterica brandenburg*	Clinical
Salmonella	*enterica derby*	Clinical
Salmonella	*enterica virchow*	Clinical
Salmonella	*enterica typhimurium*	Clinical
Salmonella	*enterica paratyphi B*	Clinical
Serratia	*liquefasciens*	ATCC 27592
Serratia	*marcescens*	ATCC 13880
Shigella	*sonnei*	Clinical
Shigella	*flexneri*	-
Vibrio	*parahaemolyticus*	-
Vibrio	*cholerae*	-
Yersinia	*pestis*	-

The following sequences have been characterized and used for multigenotypic identification of Gram-negative bacteria.

The **HI1576 sequence** (marker VI) from *Haemophilus influenzae* corresponds to the gene coding for phosphoglucose isomerase (accession n° U32831, from position 12660 to 13991) an enzyme playing a role in glucidic metabolism especially for glycolysis (Morris *et al*, 2001).

Another sequence is the **Ecs0036 sequence** (marker V) from *Escherichia coli* O157:H7 (accession n° AP002550; from position 35200 to 36200). It is believed that this sequence encodes the large carbamoyl-synthetase unit, an enzyme which catalysis the synthesis of carbamoyl phosphate, from glutamine, bicarbonate and two ATP molecules through a mechanism which requires several successive steps (Raushel *et al*, 2001). The synthesized carbamoyl-phosphate contributes to *de novo* synthesis of pyrimidic bases in bacteria (Minic *et al*, 2001).

Another sequence is the **EG10839 & EG11396 (sfrB & yigC) sequence** (= marker VII) from *Escherichia coli* K12 (accession n° NC_000913; from position 4022578 to 4024071). The corresponding protein is not yet known. When considering the sequence of the gene, it is a putative flavoprotein reductase. A search in DNA databases allowed us to find homologous sequences in some bacteria.

The **HI0019** (= marker VIII) sequence from *Haemophilus influenzae* (accession n° U32687, from position 7501 to 8550). This sequence shares all the characteristics of a coding sequence. However, the product of this coding sequence does not match any known

protein. This sequence is homologous to the gene coding for the hypothetical protein yleA from *Pasteurella multocida* (accession n° AF23940) whose function is totally unknown.

Example 3 Additional molecular marker sequences

This example illustrates the determination of marker sequences homologous to the sequences Spy0160, Spy1372, SpyM3-0902 & SpyM3-0903, Spy1527, Ecs0036, HI1576, and EG10839 & EG11396 as defined above in other Gram-positive or Gram-negative bacteria. These marker sequences were obtained using sets of degenerated primers, which have been identified based on theoretical alignments of the above-defined marker sequence (see examples 1-2) with sequences available in DNA databases.

Table 3A-H represents some degenerated primer sequences that have been used to amplify homologous marker sequences present in other Gram-positive or Gram-negative bacteria. PCR amplification was done using the identified degenerated primers on bacterial DNA isolated from various other Gram-positive or Gram-negative bacteria. Amplification was done under conditions of low stringency. In **table 3A-H**, the conditions the PCR programs and applied temperatures are indicated.

Table 3A Target sequence: Homologs of Spy0160 (purA gene or Marker I) in Gram-positive bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GRP1-S	5'-YTHTTYGAAGGDGCDCAAGG-3' (SEQ ID NO: 462)	61°C	50°C	585 bp
GRP1-AS	5'-GRYCWGGMCCWACTGAGAA-3' (SEQ ID NO: 463)	59°C		

Table 3B Target sequence: Homologs of Spy1372 (pstI gene or Marker II) in Gram-positive bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GRP2-S	5'-CCNGCCATYTCWCCRCACAT-3' (SEQ ID NO: 464)	63°C	50°C	443 bp
GRP2-AS	5'-AMGARATGAAYCCRTTCYTDGG-3' (SEQ ID NO: 465)	64°C		

Table 3C Target sequence: Homologs of SpyM3_0902 & SpyM3_0903 (Marker III) in Gram-positive bacteria

Primers	Sequence	Tm	Annealing temperature	Amplicon size

			during the PCR	
GRP3-S	5'-GACGGAMYTCTGGAGAGACC-3' (SEQ ID NO: 466)	57°C	48°C	around 600 bp
GRP3-AS	5'- GCRTAYTTDGTDGCCATWCCAAA-3' (SEQ ID NO: 467)	59°C		

Table 3D Target sequence: Homologs of Spy1527 (typA gene-Marker IV) in Gram-positive bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GRP4-S	5'-GARCGTATYATGAAAATGGT-3' (SEQ ID NO: 468)	57°C	45°C	885 bp
GRP4-AS	5'-CATDCCYTCAGDCKCAT-3' (SEQ ID NO: 469)	59°C		

Table 3E Target sequence: Homologs of HI1576 (glucose-6-phosphate isomerase gene - marker VI) in Gram-negative bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GN-1-S	5'- TGGGTYGGYGGYCGTTACT-3' (SEQ ID NO: 470)	63°C	50°C	around 500 bp
GN-1-AS	5'- TCGGTYTGNGCRAAGAAGTT-3' (SEQ ID NO: 471)	64°C		

Table 3F Target sequence: Homologs of Ecs0036 (Carb-P, large subunit gene- or Marker V) in Gram-negative bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GN-2-S	5'-CSACNATYATGACYGAYCC-3' (SEQ ID NO: 472)	63°C	50°C	500-650 bp
GN-2-AS	5'-TCCATYTCRTAYTCYTTCCA-3' (SEQ ID NO: 473)	64°C		

Table 3G Target sequence: Homologs of EG10839 & EG11396 (sfrB & yigC or Marker VII) in Gram-negative bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GN-3-S	5'- AAYTTGGTRTACATRAACTG -3' (SEQ ID NO: 474)	63°C	50°C	Around 600 bp
GN-3-AS	5'- RVTGATYATGCGYTGGCT-3' (SEQ ID NO: 475)	64°C		

Table 3H Target sequence: Homologs of HI0019 (yleA or Marker VIII) in Gram-negative bacteria

Primers	Sequence	Tm	Annealing temperature during the PCR	Amplicon size
GN-4-S	5'- GCCNGGRAADCCNACRAT -3' (SEQ ID NO: 476)	63°C	60°C	Around 500 bp
GN-4-AS	5'- GTNTCNRTNATGGAAGGCTG-3' (SEQ ID NO: 477)	64°C		

An example of a PCR amplification used to obtain marker sequences from bacteria is as follows: 10 ng of genomic DNA from each bacterial strain tested is added to a mixture containing 10 mM Tris HCl pH 9, 2.5 mM MgCl2, 50 mM KCl, 0.1% Triton X-100 (v/v), 300 nM of each primer (forward and reverse; see **table 3** for primers used), 0.25 mM desoxynucleotides triphosphates (Roche Diagnostics, Mannheim, Germany), 2.5 U de Taq Polymerase Expand High Fidelity (Roche Diagnostics, Mannheim, Germany) in a final volume of 50 µl. Amplifications were carried out in a Mastercycler gradient (Applied Biosystem 2400, USA). An initial activation step of Expand High Fidelity (94°C for 3 min) is followed by 35 cycles (94°C for 40 s, annealing temperature equal to Tm -5 or -10 °C for 50 s, 72°C for 1 min) and a final extension for 10 min.

Amplified DNA fragments were run on a 2% agarose gel stained with ethidium-bromide and visualized on a UV transilluminator.

Figures 1-3 illustrate the amplification in some Gram-positive bacteria of molecular markers which are homologous to markers I to III respectively. **Figures 8** and **10** illustrate the amplification in some Gram-negative bacteria of molecular markers homologous to markers V (Ecs0036) and VI (HI1576) respectively. **Figures 4-7, 9** and **11-21** illustrate molecular marker sequences from different Gram-positive bacteria or Gram-negative bacteria.

Example 4 Use of an assay according to the invention for molecular identification of various *Bacillus* species and strains

The present example illustrates the use of an assay according to the invention for the molecular identification of various *Bacillus* species, including *Bacillus anthracis*, species that cannot be discriminated with ribosomal 16S rDNA gene (La Scola *et al*, 2003). **Table 4** summarizes the results obtained for the various *Bacillus* species with the current genotyping assay. The analyses were performed using the markers I, II and III. Marker I corresponds to the Spy0160 sequence, Marker II corresponds to the Spy1372 sequence, Makrer III corresponds to the SpyM3_0902 & Spy M3_0903 sequence.

Primers where designed for these markers and with these primers DNA of *Bacillus* was amplified. Then, the obtained amplified sequences were sequenced and compared by alignment. **Table 4** indicates the number of modified nucleotides in the amplified marker sequences of these different *Bacillus* species and strains. The 16S ribosomal marker is not relevant in this context and cannot be used for identifying *Bacillus* species because the amplicons corresponding to each *Bacillus* species will give comparable nucleic acid sequences, with variations not exceeding 1% on the whole gene.

Table 4 Use of different molecular markers for the identification of *Bacillus* species

	Molecular marker	*B. cereus 10987*	*B. cereus 14579*	*B. thuringiensis 4Q2-72 israelensis*	*B. anthracis 1978*
B. cereus 10987	Marker III (SEQ ID NO: 168)		8	29	29
	Marker I (SEQ ID NO: 18)		1	23	21
	Marker II (SEQ ID NO: 70)		2	3	11
	16S		0	0	ND
B. cereus 14579	Marker III (SEQ ID NO: 169)	8		32	31
	Marker I (SEQ ID NO: 19)	1		22	20
	Marker II (SEQ ID NO: 71)	2		5	13
	16S	0		0	ND
B. thuringiensis 4Q2-72 israelensis	Marker III (SEQ ID NO: 170)	29	32		12
	Marker I (SEQ ID NO: 33)	23	22		30
	Marker II (SEQ ID NO: 80)	3	5		12
	16S	0	0		ND
B. anthracis 1978	Marker III (SEQ ID NO: 162)	29	31	12	
	Marker I (SEQ ID NO: 12)	21	20	30	
	Marker II (SEQ ID NO: 64)	11	13	12	
	16S	ND	ND	ND	

ND = not determined

This example shows unambiguously that the combined use of several unrelated molecular markers markedly improves bacterial species identification, as well as, to some extent, characterization of a well-determined strain within a particular species. The present

method is so specific that it can go beyond the species identification and discriminate strains of the same species.

Example 5 Nucleic acid probes according to the present invention

This example illustrates an oligonucleotide (nucleic acid probe) that has been designed from marker I of *Staphylococcus aureus* (SEQ ID NO: 23) 5'-gtgtaggtcctacattcgtttc-3' (SEQ ID NO: 478). This oligonucleotide is specific for *S. aureus* species and can therefore allow discrimination of this species with other bacteria whereas another nucleic acid probe, the oligonucleotide 5'-cattcgtttcaaaggtaatg-3' (SEQ ID NO:479) which is located on the same marker allows discrimination of different strains of *S. aureus* (i.e. strains MRSA MW2 and MRSA COL versus strains Mu-50 and N315). These multi-resistant strains carry different methicillin resistance cassette chromosome and their resistance patterns to antimicrobials agents are different.

The illustrated nucleic acid probes can for instance advantageously be used in an assay according to the present invention, on a DNA chip according to the present invention. The two oligonucleotides provided above can be considered as specific probes which can be bound on a biochip and therefore allow discrimination between various amplicons obtained from MRSA strains amplified with the primers of **table 3**.

Conclusion

The present invention demonstrates that multigenotypic molecular analyses according to Gram-, genus– species- and strain-specificity can be achieved by using concomitantly or sequentially a panel of distinct conserved molecular markers, either by conventional polymerase chain reaction PCR (with exploitation of single nucleotide specific polymorphism or SNP) , real-time PCR (with/without specific Taqman probes), or post-PCR reverse hybridization on solid support (micro-, macro- or nano-array). The analyses allow a fast and specific detection of bacterial DNA and a wide bacterial genotyping in human, animal or environmental samples.

The combined use of the herein described molecular markers allows rapid and specific molecular identification of a wide panel of bacteria in samples and/or tissues, even in samples showing a background bacterial flora. To the applicant's knowledge, there is no such diagnostic tool that is based on the use of a panel of various highly conserved bacterial moleculer markers for detecting and identifying bacteria according to Gram-, genus-, species-, and to some extent also strain- classification. Compared to the existing typing systems, we believe that this is a major improvement, in view of the increased need for rapid

and multigenotypic bacterial diagnoses, especially when considering nosocomial infections and epidemic bacterial diseases occurring in a natural, accidental or criminal setting. To this respect, molecular typing of bacteria according to the Gram phenotype is of particular interest when appropriate antibiotherapy has to be rapidly started.

References

Bosshard PP, Abels S, Altwegg M, Bottger EC, Zbinden R. Comparison of conventional and molecular methods for identification of aerobic catalase-negative gram-positive cocci in the clinical laboratory. J Clin Microbiol 2004 ;42:2065-2073.

Bosshard PP, Abels S, Zbinden R, Bottger EC, Altwegg M. Ribosomal DNA sequencing for identification of aerobic gram-positive rods in the clinical laboratory (an 18-month evaluation). J Clin Microbiol 2003;41:4134-3140.

Brakstad OG, Aasbakk K, and Maeland JA. Detection of *Staphylococcus aureus* by polymerase chain amplification of the nuc gene J Clin Microbiol. 1992;32:1768-1772.

Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004;17:840-862.

Giannino V, Santagati M, Guardo G, Cascone C, Rappazzo G, Stefani S. Conservation of the mosaic structure of the four internal transcribed spacers and localisation of the rrn operons on the Streptococcus pneumoniae genome. FEMS Microbiol Lett 2003 ;223:245-252.

Gurtler V and Stanisich VA. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. Microbiology 1996;142:3-16.

Gutierrez A, Ramos MA, Sanz JC, Bernal A, Agirrezabal J, Casado Y, Martinez M.. Bacterial meningitis in emergency medicine. Factors associated with delay of antimicrobial therapy. Enferm Infecc Microbiol Clin 1998 ;16:302-306.

Jonas D, Spitzmuller B, Weist K, Ruden H, Daschner FD. Comparison of PCR-based methods for typing Escherichia coli. Clin Microbiol Infect 2003 ;9:823-831.

Klaschik S, Lehmann LE, Raadts A, Book M, Hoeft A, Stuber F. Real-time PCR for detection and differentiation of gram-positive and gram-negative bacteria. J Clin Microbiol 2002; 40:4304-4307.

La Scola B, Zeaiter Z, Khamis A, Raoult D. Gene-sequence-based criteria for species definition in bacteriology: the Bartonella paradigm. Trends Microbiol 2003 ;11:318-321.

Lecouvet F, L.M. Irenge, B. Vandercam, A. Nzeusseu, S. Hamels, J.L. Gala. The etiologic diagnosis of infectious discitis is improved by amplification-based DNA analysis. Arthritis & Rheumatism 2004;9:2985-2994.

Palomares C, Torres MJ, Torres A, Aznar J, Palomares JC.Rapid detection and identification of Staphylococcus aureus from blood culture specimens using real-time fluorescence PCR. Diagn Microbiol Infect Dis 2003;45:183-189.

Poyart C, Quesne G, Boumaila C, and Trieu-Cuot P. Rapid and accurate species-level identification of coagulase-negative Staphylococci by using the soda gene as a target. J Clin Microbiol 2001;39:4296-4301.

Trampuz A, Steckelberg JM, Osmon DR, Cockerill FR, Hanssen AD, Patel R. Advances in the laboratory diagnosis of prosthetic joint infection. Reviews in Medical Microbiology . Reviews in medical Microbiology 2003; 14: 1-14.

Tzanakaki G, Tsolia M, Vlachou V, Theodoridou M, Pangalis A, Foustoukou M, Karpathios T, Blackwell CC, Kremastinou J. Evaluation of non-culture diagnosis of invasive meningococcal disease by polymerase chain reaction (PCR). FEMS Immunology and Medical Microbiology 2003; 39 : 31-36.

Xu J, Rao JR, Millar BC, Elborn JS, Evans J, Barr JG, Moore JE. Improved molecular identification of Thermoactinomyces spp. associated with mushroom worker's lung by 16S rDNA sequence typing. J Med Microbiol 2002;51:1117-1127.

Yu DT, Black E, Sands KE, Schwartz JS, Hibberd PL, Graman PS, Lanken PN, Kahn KL, Snydman DR, Parsonnet J, Moore R, Platt R, Bates DW. Severe sepsis: variation in resource and therapeutic modality use among academic centers. Crit Care. 2003;7:R24-R34.